

TOPIK 말하기 평가 기준의 재설정을 통한
화용 능력 평가 연구
— 2번 ‘그림 보고 역할 수행하기’, 4번 ‘대화 완성하기’를
중심으로

성인경 동명대학교 글로벌한국학과 초빙교수

* 이 논문은 2024년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구
(NRF-2024S1A5B5A17034494)이며, 제85회 국어교육학회 학술발표대회 (2025.8.23.)
에서 발표한 것을 수정·보완한 것이다.

- I. 서론
- II. 이론적 배경 및 선행 연구
- III. 연구 대상 및 연구 방법
- IV. 연구 결과
- V. 결론 및 제언

I. 서론

이 연구는 현재 실시되고 있는 TOPIK(Test Of Proficiency in Korean: 한국어 능력 시험) 말하기 평가의 기준이 한국어 학습자의 한국어 말하기 능력을 평가하기에 적합한가를 고찰하고, 화용 능력 평가의 관점에서 2번, 4번 문항의 평가 기준을 재설정하여 제안하고자 한다.

TOPIK 말하기 평가(이하 토픽 말하기 평가)의 목적은 한국어를 모국어로 하지 않는 재외 동포 및 외국인에게 의사소통 중심의 한국어 학습 방향을 제시하고, 이들의 의사소통 능력을 측정하고 평가하여 그 결과를 국내 대학 유학 및 취업에 활용하고자 하는 것이다.¹⁾ 기존의 토픽 시험은 듣기, 읽기, 쓰기만을 평가하므로 종합적인 의사소통 능력을 평가하기 위한 한국어 말하기 평가의 필요성이 제기되어 왔고 현재 8회까지 실시되었다(2025년 10월 18일 기준).

토픽 말하기 평가의 목적이 의사소통 중심의 한국어 학습을 제시하고

1) 토픽 한국어 능력시험 홈페이지(www.topik.go.kr) 참고.

한국어 학습자의 의사소통 능력 전반을 평가하고자 하는 것이라면 평가 문항과 평가 기준도 그 목적에 맞게 타당하게 구성되어야 한다. 현재 토픽 말하기 평가의 유형은 6가지로 다양하게 구성되어 있으며 각 유형이 측정하는 수준 역시 초급, 중급, 고급으로 구별되어 설정되어 있다. 그러나 평가 기준은 1~6번 문항이 모두 동일하며 평가 기준에 대한 정보가 구체적으로 제공되지 않아 교수자와 학습자 모두 토픽 말하기 시험에 접근하기가 쉽지 않은 것이 사실이다.

평가 기준은 교육 방법에 긍정적, 부정적 영향을 미치는 환류효과(washback-effect)를 가진다. 즉, 말하기 평가의 기준이 교수자가 학습자에게 중점적으로 가르치고자 하는 내용에 큰 영향을 주므로 학습자들 역시 학습의 방향을 평가 기준에 맞추게 된다는 것이다. 그런데 평가 기준이 평가의 목적에 맞지 않거나 학습자의 능력을 타당하게 평가하지 못할 경우, 학습자는 평가 기준을 만족시키는 데에만 집중하게 되어 교육 목표에 도달하기 어려워지는 문제가 발생한다.

토픽 말하기 평가가 한국어 학습자에게 의사소통 중심의 한국어 학습을 제시하는 도구로 기능하기 위해서는 의사소통의 맥락과 상황에 긴밀하게 관련되는 ‘화용능력(pragmatic competence)’까지도 종합적으로 평가할 수 있어야 한다고 본다. 화용능력은 의사소통 능력을 이루는 중요한 능력 중 하나로 실제 언어 상황에서 상황과 맥락을 잘 이해하고 적절하게 언어를 사용할 수 있는 능력이다.

따라서 이 연구에서는 기존의 토픽 말하기 평가 기준을 비판적으로 고찰하여 이를 토대로 한국어 말하기 화용 능력을 포함한 전반적인 의사소통 능력을 평가할 수 있는 평가 기준으로서 재설정하기 위한 방향을 제시하고자 한다.

II. 이론적 배경 및 선행 연구

1. 토픽 말하기 평가 연구

토픽 말하기 평가가 실시된 2022년 이후 토픽 말하기 평가 연구는 소논문 14편, 석사 논문 1편, 박사 논문 2편 정도이다.²⁾ 연구 분야는 말하기 평가의 특정 문항에 대한 교육 방법(조희영·김지현, 2023; 조희영, 2024; 김도연·난춘야·문보라, 2023), 토픽 말하기 평가의 특징에 대한 분석(양세정, 2023)이나 토픽 말하기 시험에 대한 학습자의 인식(위약요, 2023)에 대한 연구 등이다. 최근에는 생성형 AI를 활용하거나 메타버스를 활용하여 토픽 말하기 학습 및 연습 활동을 개발한 연구(위약요·최은경·김성주, 2023; 장지영·박수연, 2025)가 활발하며 ChatGPT를 학습의 피드백 도구로 활용하여 토픽 말하기를 연습하거나 피드백 모형을 개발한 연구(김유미, 2024) 등도 이루어지고 있다.

조희영·김지현(2023)에서는 토픽 말하기 평가 5, 6번 문항에서 학습자가 어려워하는 부분을 확인하고 효과적인 교수·학습 방안을 제시하였고, 조희영(2024)에서는 4번 대화 완성하기 문항을 모의 평가 한 결과로 문항 교육 방안을 제시하고 있다. 김도연·난춘야·문보라(2023)에서는 토픽 말하기 평가 시범 시험에 참여한 고급 학습자를 대상으로 인터뷰를 진행하고, 오류 양상을 분석하여 교육방안을 제시하였다.

토픽 말하기 평가는 기존에 시행되어 오던 토픽 시험에 비해 응시해 본 학습자가 적으며 IBT 방식에 익숙하지 않아 시험 상황에서 긴장하는 경우가 많다. 이에 양세정(2023)은 토픽 말하기 평가의 형식이나 유형, 모범답안 및

2) 학술연구정보서비스(www.riss.kr), 한국학술지인용색인(www.kci.go.kr)에서 ‘토픽 말하기 평가’를 키워드로 검색하였다.(2025년 10월 18일 검색 기준)

응시자들의 피드백을 분석하여 제시하였고, 위약요(2023)에서도 토픽 말하기 평가에 응시한 중국인 학습자의 불안감을 파악하고 이를 낮추기 위한 교육 현장에서의 교육 방향을 모색하고 있다.

한편 장지영·박수연(2025)에서는 멀티모달 기반의 음성 상호작용이 가능한 생성형 AI를 활용하여 학습자들이 토픽 말하기 평가를 실시간으로 연습하고 자율 피드백을 받을 수 있는 도구로서 활용 가능성을 보여주고 있다. 이와 같은 맥락에서 김유미(2024)에서도 ChatGPT의 즉각성과 맞춤형 피드백을 활용하여 토픽 말하기를 연습하고 학습자들의 토픽 말하기 평가에 대한 불안과 어려움을 다소 해결해 줄 수 있는 방안을 제시하고 있다. 이처럼 토픽 말하기 평가와 관련된 선행 연구들은 토픽 말하기 시험의 특성에 대한 파악이나 교육 방안을 중심으로 이루어져 왔고, 토픽 말하기 평가 기준에 대한 연구는 이은영(2024)이 유일하다. 이은영(2024)은 토픽 말하기 평가를 주제로 한 박사 논문으로 토픽 말하기 평가의 구인의 타당도를 높이기 위한 방안을 설계하였다. 이 연구에서는 ‘과제 수행 능력, 담화 구성 능력, 어휘 및 문법 능력, 유창성 및 발음 능력, 사회언어학적 능력, 상호 작용 능력’의 6가지 평가 구인을 설정하고 초급, 중급, 고급 단계에서 각 평가 구인의 구현율을 높이기 위해 문항 유형과 문항 수도 늘리는 개선안을 마련하였다. 이 연구는 화용 능력을 평가 기준으로 삼지는 않았으나 토픽 말하기 평가가 의사소통 능력을 구성하는 다양한 능력을 타당하게 평가하는 도구로 기능하도록 하는 기초 연구로서 의의가 있다고 본다.

2. 화용 능력 평가 연구

화용 능력은 앞서 밝힌 것처럼 의사소통의 상황과 맥락에 관련된 능력이다. 이 연구에서는 한국어 말하기 화용 능력을 ‘한국어 의사소통 상황에서 상황과 맥락에 맞게 언어를 적절하게 사용할 수 있는 능력’(성인경¹, 2024: 5)으로 정의하고자 한다. 화용 능력에 대한 연구들은 주로 화용전략과 화행

을 중심으로 이루어져 왔으며, 화용적 성공과 실패에 대한 연구 및 화용 능력 향상을 위한 교육 방안 등 화용 능력을 ‘교육의 대상’으로 삼은 연구가 대부분이었다. 그러나 최근 한국어 교육 분야에서 화용 능력을 ‘평가의 대상’으로 삼아 화용 능력 평가 방법 및 기준, 화용 능력의 구인 등에 대한 연구들이 이루어지고 있다.

화용 능력은 이해와 표현 측면에서 모두 요구되는데 강수정·김영규(2016)에서는 이해의 측면에서 한국어 능력시험의 읽기 영역을 대상으로 평가의 구인을 설정하여 제시하였다. 이 연구에서는 화용 능력을 논리적 추론, 화용적 추론, 사회 화용론, 화용 언어로 나누어 측정해야 하며 평가의 구성 요소를 다양하게 제시하고 있다. 그러나 화용능력 평가의 구인을 밝히고, 평가 요소를 제시할 뿐 실제 평가의 상황에서 어떻게 적용시켜야 할지 구체적인 적용이 이루어지지 않았다. 정민수(2018)에서는 화용 능력을 사회화용적 능력, 화용언어적 능력으로 구성되어 있다고 보고 이를 평가하기 위한 선다형 문항 및 적절성 판단 테스트 등을 개발하여 제시하고 있으나 이 역시 이해 측면에서의 화용 능력 평가 도구이다.

강현주(2018)에서는 한국어 말하기 평가 맥락에서 사회언어적 능력을 발화 맥락에 따라 발화의 내용과 형식을 변화할 수 있는 능력으로 보고, 이를 보조화행, 호칭어, 높임법을 사용하여 측정하는 방안을 제안하였다. 이 연구에서는 사과 화행을 대상으로 삼아 사회언어학적 능력이 어떻게 측정될 수 있는지를 구체적으로 제시하고 있다.

그러나 화용 능력의 평가 기준이나 구인에 대한 연구는 활발히 이루어지지 못했다. 연구자에 따라 화용 능력을 말하기 평가의 구인에 포함시키지 않는 경우도 있고, 말하기 평가의 구인에 포함시키더라도 화용 능력을 평가하기 위한 기준에 대한 연구는 찾아보기가 힘들다. Ishihara·Cohen(2018)은 화용 능력 평가의 다양한 방법들을 소개하고 평가의 기준을 담화의 성격에 따라 평가지 형식으로 제시하고 있다. 이 연구에서는 화용능력을 화용언어적 측면, 사회화용적 측면, 메타화용적 측면으로 나누어 평가할 수 있다고

보고, 실제 평가에 적용할 수 있는 다양한 루브릭 평가지와 평가 기준을 예로 보여주고 있으나 평가의 기준이 일률적이지 않아 적용에 어려움이 있다.

성인경(2024 ㄱ)에서는 한국어 학습자의 화용 능력 평가를 하기 위한 평가의 구인을 화행을 대상으로 하여 밝히고 있다. 이 연구에서는 화용 능력 평가의 구인으로 문화적 요인과 언어적 요인으로 나누고 문화적 요인에 화자와 청자의 관계적 요소, 사회적 요소를 언어적 요인에 구성적 요소와 표현적 요소로 나누어 평가할 수 있다고 하였다. 성인경(2024 ㄴ)에서는 이러한 평가의 구인을 토대로 실제 화용 능력 평가 도구를 평가 루브릭으로 구현하여 제시하고 있다. 화용 능력 평가의 구인을 바탕으로 실제 평가의 기준을 세우고 개발한 루브릭을 실제 평가 상황에 적용하였다.

그러나 화용 능력 평가는 화행과 같은 특정한 담화 유형을 평가하기 위한 기준일뿐만 아니라 의사소통 능력 평가의 중요한 기준이 된다. 따라서 토픽 말하기와 같은 공신력 있는 평가 도구의 평가 기준을 화용 능력 평가의 관점에서 고찰하고, 토픽 말하기 평가가 학습자의 의사소통능력 전반을 타당하게 평가할 수 있는 평가도구로 기능할 수 있도록 평가 기준을 재설정하는 것이 필요하다고 본다.

III. 연구 대상 및 연구 방법

1. 연구 대상

토픽 말하기 평가의 문항은 총 6개의 유형으로 이루어져 있고, 한국어 학습자의 숙달도에 따라 측정 범위를 다르게 설정하였다. 그러나 토픽 말하기 평가의 요소는 다음의 <표 1>과 같이 1~6 문항 모두 동일하게 제시되어 있다.

〈표 1〉 토픽 말하기 평가 요소

평가 요소	내용
내용 및 과제 수행	과제에 적절한 내용으로 표현하였는가? 주어진 과제를 풍부하고 충실하게 수행하였는가? 담화 구성이 조직적으로 잘 이루어졌는가?
언어사용	담화 상황에 적합한 언어를 사용하였는가? 어휘와 표현을 다양하고 풍부하게 사용하였는가? 어휘와 표현을 정확하게 구사하였는가?
발화 전달력	발음과 억양이 어느 정도 이해 가능한가? 발화 속도가 자연스러운가?

위의 표에서 보듯 토픽 말하기 평가에서는 ‘내용 및 과제 수행, 언어사용, 발화 전달력’의 세 가지 요소를 평가하고 있다. 그러나 이 평가 요소가 정확히 어떤 의미인지에 대한 설명이 부족하고, 각 문항의 배점에서 이 3가지 요소가 차지하는 배점도 제시되지 않아 평가 기준에 대한 정보가 매우 부족한 편이다.

특히 토픽 말하기 2번과 4번은 다른 유형들에 비해 제시되는 상황과 청자를 고려하여 말하는 능력이 필수적인 유형이라고 볼 수 있다. 2번 문항은 길 찾기, 물건 사거나 교환하기, 공공시설 이용하기 등 일상생활에서 수행하는 언어 기능에 대한 평가 유형이고, 4번 문항은 상대방의 말을 듣고 그에 적절하게 대응하여 대화를 완성하는 유형이다. ‘화용 능력’은 발화의 상황과 맥락을 고려하여 적절하게 이해하고 표현할 수 있는가에 관련된 능력이므로, 다른 문항에 비해 2번과 4번 문항이 이 연구에 적합한 문항이라 할 수 있다.³⁾

- 3) 1번 문항은 일상생활과 관련된 짧은 질문을 듣고 말하는 유형으로, 일방적인 말하기이며 상대방과의 관계나 상황과 관련이 없는 말하기이다. 3번은 그림을 보고 순서대로 이야기를 만드는 문항으로 화용 능력보다는 그림의 내용을 충실하게 반영하고, 연결하는 말을 잘 사용하여 말할 수 있는 능력이 중요하다. 5번 문항은 자료를 보고 해석하여 비판적으로 자신의 의견을 말하는 능력이 요구되며 6번 문항 역시 자신의 견해를 논리적으로 말하는 문항으로 5번, 6번은 논리적인 내용 구성과 정확한 언어 사용이 더 중요한 문항이라 할 수 있다. 넓은 의미에서는 1~6번 문항 모두 화용 능력이 요구되기는 하나, 이 연구에서는 2번과 같이 상황 맥락을 고려하여 내용을 달리 구성해야 하는 문항, 4번과 같이 대화 상대

이 연구의 대상이 되는 문항은 다음과 같다.

2번 문항	4번 문항
약속 시간에 늦을 것 같습니다. 친구에게 전화해서 늦는 이유를 이야기해 보세요.  여자: 어보세요? 왜 아직도 안 와?	두 사람이 중간 소음에 대해 이야기하고 있습니다. 여자의 마지막 말을 듣고 남자가 할 말로 대화를 완성해 보세요.  여자: 민수 씨, 저 요즘 힘든 일이 있어요. 남자: 수민 씨, 무슨 일이에요? 그리고 보니 기운이 없어 보여요. 여자: 얼마 전에 댁집이 새로 이사 왔는데 중간 소음 때문에 밤에 잠을 잘 못 지요. 애들이니까 이해해 보려고 해도 계속 반복 되니까 너무 힘들어요.

〈그림1〉 연구 대상 2번, 4번 문항⁴⁾

〈그림 1〉과 같이 토픽 말하기 시험과 동일한 유형의 문항을 제시하여 학습자 6인이 말하기를 수행하고, 그 결과를 토픽 말하기 평가의 기준으로 평가한 뒤 결과를 분석할 것이다.

2. 연구 방법

이 연구의 목적인 토픽 말하기 평가 기준 재설정을 위해 우선 한국어 학습자의 말하기 수행 자료를 녹음하였다. 2025년 5월 8일 중급 이상의 한국어 학습자 6인⁵⁾을 대상으로 토픽 말하기 평가 2번과 4번 유형의 말하기를 수

자와의 관계 및 주제를 고려하여 말하기 내용 및 표현을 조정해야 하는 문항에서 화용 능력을 보다 중요하게 평가해야 한다는 관점에서 연구 대상을 2번, 4번 문항으로 선정하였음을 밝힌다.

4) 이 문항은 COOL TOPIK 말하기(2022)에 제시된 예시 문항을 사용하였다.
5) 한국어 학습자 6인에 대한 정보

	한국어 능력 수준	한국 학습 기간	국적
S1	토픽 4급	4년 반	우즈베키스탄

행하게 하였다. 말하기 수행 시, 준비 시간과 수행 시간을 토픽 말하기 평가와 동일하게 설정하였고 평가 점수 역시 동일하게 부여하였다.⁶⁾

학습자 수행 자료는 녹음 파일과 함께 클로바노트(clova note)로 전사하여 평가 전문가⁷⁾에게 제공하였다. 녹음 파일을 듣고 평가하되, 전사 자료를 참고할 수 있도록 하였고 수행 자료는 2번 듣고 평가하도록 동일하였다. 1차 평가는 2번과 4번의 문항을 현재의 평가의 기준으로 평가하면서 평가 기준의 타당성과 평가 시 발생하는 어려움을 파악하기 위한 것이다. 1차 채점 결과에 대한 전문가 집단의 의견을 수렴한 뒤 기존의 토픽 말하기 평가 항목을 수정하였는데, 수정 과정에서 전문가 집단 간 합의가 필요한 부분은 의견을 수렴하여 전문가 집단과 함께 2차 평가를 위한 평가 기준을 만들었다.

그 후 수정된 평가 기준으로 동일한 학습자 수행 자료를 2차로 평가하였는데, 동일한 학습자의 수행이므로 채점 결과가 학습되는 것을 방지하기

S2	토픽 6급	6년	베트남
S3	토픽 6급	4년	미얀마
S4	토픽 6급	4년 반	미얀마
S5	토픽 5급	3년	베트남
S6	토픽 4급	2년	베트남

6)

문항		준비 시간	수행 시간	배점
2번	그림 보고 역할 수행하기	30초	40초	9점
4번	대화 완성하기	40초	60초	12점

7) 전문가 집단은 한국어 교육 전문가이자 말하기 평가 전문가 5인으로 구성하였다. 평가자 5인 중 4인이 한국어 교육학 박사학위 소지자이며, 1인은 박사 수료자이다. 전문가 집단의 평균 한국어 교육 경력은 8년 이상이며 5인 중 2인은 토픽 말하기 교재 집필진이며, 토픽 말하기의 교육 방법, 한국어 발음 채점 방법 등 말하기 교육 및 평가와 관련된 다수의 논문을 집필하였다. 또한 평가자들은 현재 언어교육원, 학부, 대학원 등에서 활발히 한국어 교육을 실시하고 있어 한국어 학습자의 말하기 평가 경험이 풍부한 말하기 평가 전문가 집단이라 볼 수 있다.

위해 1차 채점과 2차 채점 사이의 기간을 두 달 정도 두고 채점을 실시하였다. 1차 채점은 2025년 5월 25 ~ 5월 29일, 2차 채점은 2025년 7월 22일 ~ 7월 25일에 실시하였다. 1차 채점과 2차 채점 사이에 전문가 집단 간의 평가 기준 수정에 대한 의견교환이 있었다. 그 후 최종 평가 기준과 평가 루브릭을 마련하고, 2차 채점을 한 후 최종 평가 기준에 대한 의견을 제시하도록 하였다. 이상의 연구 과정을 보이면 다음의 <그림 2>와 같다.



<그림 2> 연구 절차

IV. 연구 결과

1. 평가 기준의 1차 수정

기존의 토픽 말하기 평가 기준을 검토하기 위해서 전문가들은 1차로 학습자 수행 자료를 평가하였고, 평가 기준의 수정 사항에 대해 2번 문항과 4번 문항으로 나누어 의견을 제시하였다. 1차 채점 후, 기존의 토픽 말하기 평가 기준에 대한 평가자들의 의견에서 ‘억양과 뉘앙스, 사회적 사용역과 맥락

에 맞지 않는 발화, 사회문화적 맥락’ 등 화용 능력과 관계된 의견이 많았다. 5인의 평가자가 공통적으로 제시한 의견도 있었고, 평가자마다 다른 개별적인 의견도 있었는데 이를 정리하면 다음과 같다.

공통 의견

1. 각 항목별 중요도에 따른 세부 배점이 제시되어야 함.
2. 언어사용 요소에서 ‘어휘와 표현을 다양하고 풍부하게 사용하였는가?’를 판단하는 것이 어려움.

개별 의견

1. 억양과 뉘앙스를 평가할 수 있는 기준이 필요함.(R2)
2. 사회적 사용역에 대한 평가 기준이 필요함.(R3)
3. 발화전달력 요소에 지나친 간투사나 말하기 지연에 대한 평가 항목이 필요함.(R5)
4. 각 문항의 성격에 따른 구체적인 평가 가이드 라인이 제공되면 좋겠음.(R1)
5. 상황 맥락에 맞지 않는 대답을 한 경우, ‘내용 및 수행’ 요소에서 평가할지, ‘언어 사용’영역에서 평가할지 헷갈림.(R5)
6. 한국의 사회문화적 맥락에 대한 이해를 평가할 수 있는 기준이 있으면 좋겠음.(R3)
7. 발화 길이에 따른 평가 가이드 라인이 제시되어야 함.(R4)

이 중에서 평가 기준과 관련된 것들은 1차 수정안에 최대한 반영하고, 발화 길이에 따른 평가 가이드 라인과 같은 평가 지침에 관련된 것은 2차 평가 전에 모든 채점자에게 전달하여 평가의 신뢰도를 높이고자 하였다. 이 의견들을 반영한 2번 문항 평가 기준을 수정한 내용은 다음과 같다.

〈표 2〉 2번 문항 평가 기준의 1차 수정안

평가 요소	내용
내용 및 과제 수행 (3-2-1-0점)	① 과제에 적절한 내용으로 표현하였는가? (내용의 적절성-기능수행 목적 달성 측면) ② 과제가 상황과 맥락에 맞게 표현되었는가? (내용의 적절성- 상황과 맥락 측면, 즉 사과, 요청, 거절 등의 상황의 성격에 맞는가?)
언어사용 (3-2-1-0점)	① 담화 상황에 적합한 언어를 사용하였는가? (표현의 적절성-높임말 등, 상황과 맥락 측면) ② 어휘와 표현을 정확하게 구사하였는가? (표현의 정확성)
발화 전달력 (3-2-1-0점)	① 발음과 억양이 어느 정도 이해 가능한가? ② 발화 속도가 자연스러운가?(말하기 지연, 간투사) ③ 말하기 불안 또는 긴장이 있는가?

1차 수정안에서 수정된 부분은 먼저 평가자들이 공통적으로 제기한 각 항목별 세부 배점을 제시한 것이다. 1차 수정안에서는 3개의 평가 요소에 동일하게 3점씩 설정하였고, 단계별 3-2-1-0의 점수를 줄 수 있도록 하였다. 그리고 내용 과제 수행, 언어사용 요소에서 기존 말하기 평가 기준은 3개의 내용으로 되어 있었으나 1차 수정안에서는 2개의 내용으로 축소하였다. 내용 및 과제 수행 요소에서는 기존 평가 기준에서 ‘주어진 과제를 풍부하고 충실하게 수행하였는가?’와 ‘담화 구성이 조직적으로 잘 이루어졌는가?’를 제외시키고 ‘과제가 상황과 맥락에 맞게 표현되었는가?’를 추가하였다. 2번 문항은 그림을 보며 질문에 대한 적절한 역할을 수행해야 하는 초급 문제이므로 과제 수행의 풍부성이나 담화 구성의 조직 보다는 과제가 상황과 맥락에 맞게 수행되었는지에 대한 적절성이 더 적합한 평가 기준이라 할 수 있다.

그리고 언어 사용요소에서는 초급 수준의 문항에서 언어 사용이 다양하고 풍부하게 이루어지기 어려우므로 이 항목을 제외하고 표현의 적절성과 정확성을 평가하는 것으로 수정하였다. 발화 전달력 요소에서는 기존 평가 기준에서 한 가지 평가 내용을 추가하여 말하기 불안 및 긴장이 있는가를 평

가하고자 하였다. 세 가지 평가 요소의 각 내용마다 그것이 무엇을 의미하는지 평가자들이 구체적으로 인식하도록 괄호로 평가 내용에 대한 설명을 덧붙였다.

평가에 대한 구체적인 가이드라인이 부족하여 평가에 어려움이 겪는다는 평가자들의 의견을 반영하여 수정된 평가 기준에 따른 평가 루브릭을 다음과 같이 개발하여 평가 신뢰도를 높이하고자 하였다.⁸⁾

〈표 3〉 2번 문항 평가 루브릭

평가 요소	점수	내용
내용 및 과제 수행	3	과제에 적절한 내용으로 표현되고, 상황과 맥락에 맞는 내용으로 대답함.
	2	- 과제에 적절한 내용이지만, 상황과 맥락에 다소 적절하지 않은 내용이 있음. - 과제에 다소 적절하지 않은 내용이 있지만, 상황과 맥락에 적절한 내용임.
	1	- 과제에 적절하지 않은 내용이 다소 있고, 상황과 맥락에도 적절하지 않은 부분이 있음. - 과제에 적절한 내용이지만, 상황과 맥락에 전혀 맞지 않는 내용으로 대답함.
	0	내용이 과제에 전혀 적절하지 않고, 상황과 맥락에도 전혀 적절하지 않음.

이 평가 루브릭은 말하기 평가 기준에 따라 각 해당 점수대에 해당되는 수준을 구체화시켜 한눈에 볼 수 있도록 한 평가표로 짧은 시간 안에 수행 자료를 듣고 평가해야 하는 말하기 평가에서 평가에 용이성을 더해주고, 평가의 신뢰성을 높일 수 있도록 하는 장치이다.

4번 문항은 대화의 맥락에 맞게 상대의 말에 적절하게 대응하는 말하기로 중급 수준의 문제이다. 따라서 2번 문항과 달리 과제 수행의 풍부성이나 담화가 조직적으로 구성되어 있는가 등에 대한 내용을 그대로 유지하였고, 평가자들이 각 평가 내용이 의미하는 바가 무엇인지를 구체적으로 인식할 수 있도록 각 평가 내용에 대한 설명을 추가하였다. 1차 수정안은 다음과 같다.

8) 평가자들에게는 '내용 및 과제 수행, 언어 사용, 발화 전달력'의 모든 내용이 포함된 루브릭을 제공하였으나 지면 관계상 일부만 제시하였다.

〈표 4〉 4번 문항 평가 기준의 1차 수정안

평가 요소	내용
내용 및 과제 수행 (4-3-2-1-0)	① 과제에 적절한 내용으로 표현하였는가? (과제 수행의 적절성) ② 주어진 과제를 풍부하고 충실하게 수행하였는가?(과제 수행의 풍부성) ③ 담화 구성이 조직적으로 잘 이루어졌는가?
언어사용 (4-3-2-1-0)	① 담화 상황에 적합한 언어를 사용하였는가? (사회문화적 맥락에 대한 이해·표현의 적절성) ② 어휘와 표현을 다양하고 풍부하게 사용하였는가?(표현의 풍부성) ③ 어휘와 표현을 정확하게 구사하였는가? (표현의 정확성)
발화 전달력 (4-3-2-1-0)	① 발음과 억양이 어느 정도 이해 가능한가? ② 발화 속도가 자연스러운가? (말하기 지연, 간투사 사용) ③ 말하기 불안 또는 긴장이 있는가?

2번 문항과 같이 총 12점을 평가 요소당 4점씩 부여하였고, 내용 및 과제 수행과 언어사용은 기존 말하기 평가와 동일하다. 발화전달력은 말하기 불안 또는 긴장에 대한 평가 내용을 추가하였다. 4번 문항 또한 수정된 평가 기준에 따라 평가 루브릭을 다음과 같이 작성하여 제공하였다.

〈표 5〉 4번 문항 평가 루브릭

평가 요소	점수	내용
내용 및 과제 수행	4	과제 수행의 적절성과 풍부성이 있고, 구성이 조직적으로 이루어짐.
	3	- 과제 수행은 다소 적절성이 떨어지나 내용이 풍부하고, 구성이 조직적임. - 과제 수행은 적절하나, 내용의 풍부성이 떨어지거나 구성의 조직성이 다소 부족함.
	2	- 과제 수행의 적절성이 다소 떨어지고, 내용의 풍부성이나 구성이 부족함. - 과제 수행은 적절하나 내용이 지나치게 간단하고 구성도 부족함.
	1	- 과제 수행의 적절성이 떨어지고, 내용도 풍부하지 못하며 구성도 매우 부족함.
	0	- 과제 수행이 적절하지 못하고, 구성이 매우 부족함.

2. 평가 기준의 2차 수정 및 최종 평가 기준

1차로 수정한 평가 기준과 루브릭을 5인의 평가자에게 제공하고 이에 대한 추가 의견을 제시하도록 하였다. 이것은 평가 기준에 대한 평가자 합의를 이루는 과정으로, 공통의견과 개별의견을 반영하여 1차 수정안을 만들었으나 추가적으로 생기는 의문이나 실제 평가에 적용 시 발생하는 문제에 대한 의견을 제시한 것이다. 1차 수정안에 대해 평가자들이 공통적으로 제시한 의견과 개별적인 의견을 정리하면 다음과 같다.

공통의견

1. 각 항목별 중요도에 따라 배점이 재조정되어야 함.
2. 말하기 불안이나 긴장을 평가하기에 어려움이 있음.

개별의견

1. 2번 문항의 내용 및 과제 수행 영역에서 '내용의 풍부성'에 대해서도 평가하면 좋겠음.(R4)
2. 2번 문항에서 과제에 적절한 내용과 상황 맥락에 맞는 내용이 어떤 차이가 있는지 설명이 필요함.(R2)
3. 답화 상황에 맞는 내용과 답화 상황에 적합한 언어 사용을 다르게 평가해야 함. 이중으로 평가될 위험이 있음.(R3)
4. 평가 시, 점수 구간이 설정되어 있는 것인지 각 평가의 요소에 맞는 점수를 구간 설정없이 부여할 수 있는지에 대한 합의가 필요함.(R2)
5. 노력을 기울여도 발음이나 억양을 이해하기 어려운 경우, 다른 영역도 모두 0 점 처리를 할 것인지에 대한 합의가 필요함.(R1)

4번 문항 보다는 2번 문항에 대한 의견이 많았고, 개별의견 4,5번은 평가 지침의 통일에 대한 의견이었다. 이 의견들을 수렴하여 2차 수정안에 반영하기 위해서 먼저 공통의견 1번 점수 배점에 관한 사항은 1차 수정안에서

각 평가요소마다 동일하게 부여한 점수를 차등적으로 부여하는 것으로 수정하였다. 내용 및 과제 수행 > 언어 사용 > 발화 전달력 순으로 1점씩 더 부여하여 2차 수정안을 작성하였다. 공통의견 2번의 말하기 불안이나 긴장의 평가에 대한 것은 시험 상황에서 긴장될 수 있고, 불안이라는 요소를 평가의 대상으로 삼는 것이 합당한가에 대한 의견이 있었다. 따라서 말하기 불안 평가 항목을 삭제하고, 대신 관찰 가능한 ‘말하기 지연이나 간투사, 반복’ 등의 항목을 ‘발화 전달력’항목에서 ‘발화 속도가 자연스러운가?’ 부분에 추가하여 평가하는 것으로 수정하였다.

개별 의견에서 2번 문항의 내용 및 과제 수행 영역에 ‘내용의 풍부성’을 평가하면 좋겠다는 의견이 있었다. 초급 문항이지만, 고급 수준의 학습자들이 풍부하게 이야기하는 경우 내용의 적절성만으로 점수를 부여하기가 어렵다는 이유이다. 따라서 기존의 말하기 평가 기준에 있었던 내용의 풍부성 기준을 추가하여 재설정하였다.

개별 의견 2번 역시 2번 문항에 관련된 내용이었는데, 과제에 적절한 내용과 상황 맥락에 맞는 내용에 어떤 차이가 있는지에 대해서는 평가자들에게 구체적인 예를 들어 설명을 추가하였다. 즉, 2번 문항이 약속에 늦은 상황에서 친구에게 약속에 늦는 이유에 대해 설명을 해야 하는 과제인데 이때 그림에 제시된 도로 공사를 포함하여 이유를 잘 설명할 경우 과제에 적절한 내용이 되고, 상황 맥락은 시간 약속에 늦은 상황이므로 사과의 내용이 포함되어야 한다는 것이다. 즉, 도로 공사 이야기는 포함되었지만 사과의 내용없이 친구에게 30분을 더 기다리라고 발화한 경우 과제에는 적절하지만 상황 맥락에는 적절하지 않은 내용이라고 할 수 있다.

개별 의견 3번도 담화 상황에 관한 것으로 담화 상황에 적절한 내용과 담화 상황에 적절한 언어 표현을 다르게 평가해야 한다는 것이다. 이에 대해서도 2차 수정안과 함께 구체적인 예시를 평가자에게 제공하였다. 즉, 시간 약속에 늦은 상황이라면 사과하는 내용이 포함되어 있는 것은 담화 상황에 맞는 내용으로 ‘내용 및 과제 수행’에서 평가되어야 하고, 담화 상황에 적합

한 언어 표현은 ‘언어 사용’ 요소에서 평가되어야 하는 것으로 사과와 관련된 어휘 선택이나 높임말 여부 등을 말한다.

그리고 개별 의견 4,5번은 평가 지침에 대한 것으로 4번 점수 구간의 설정에 대해서는 기존 토픽 쓰기 평가에서도 점수 구간 설정을 하므로 말하기에서도 점수 구간을 설정하는 것으로 안내하였다. 즉, 내용 및 과제 수행의 최고점이 3점이라면 언어 사용 및 발화전달력도 그 점수의 ± 1 점 내에서 처리하는 것으로 하였다. 이러한 합의의 과정을 통해 <표 6>과 같이 2번 문항 평가 기준을 수정하여 최종 평가 기준을 마련하였다.⁹⁾

<표 6> 2번 문항 최종 평가 기준

평가 요소	내용
내용 및 과제 수행 (4-3-2-1-0)	① 과제에 적절한 내용으로 풍부하게 표현하였는가? (내용의 적절성-기능수행 목적 달성 측면) ② 과제가 상황과 맥락에 맞게 표현되었는가? (내용의 적절성- 상황과 맥락 측면, 즉 사과, 요청, 거절 등의 상황의 성격에 맞는가?)
언어사용 (3-2-1-0)	① 담화 상황에 적합한 언어를 사용하였는가? (표현의 적절성- 높임말 등, 상황과 맥락 측면) ② 어휘와 표현을 정확하게 구사하였는가? (표현의 정확성)
발화 전달력 (2-1-0)	① 발음과 억양이 어느 정도 이해 가능한가? ② 발화 속도가 자연스러운가?

9) 표에서 볼드처리한 부분이 1차 수정안에 비해 추가되거나 변경된 부분이다. 최종 평가 기준에 따라 평가 루브릭의 내용도 수정하였으나 지면 관계상 생략하였음을 밝힌다. 또한 이 연구에서 제시하는 토픽 평가 기준은 2번, 4번에만 적용되는 것이므로 실제 토픽 말하기 평가 상황에 적용하기 위해서는 채점자 교육 등의 논의가 필요할 것이라고 본다. 다만, 이 연구에서는 평가 요소는 기존의 형식으로 유지하되, 각 문항의 성격에 따라 평가 내용은 달리 적용하는 것이 타당성 있는 평가가 될 것이라고 보았으므로 2번과 4번 문항에 적용할 수 있는 평가 내용을 재설정함으로써 토픽 말하기 평가의 타당도를 높이는 데에 기여할 수 있다고 본다.

4번 문항은 2번 문항에 비해 수정 사항이 많지 않았고, 각 평가 요소의 중요도에 따른 점수 배점의 조정과 발화 전달력 요소에서 말하기 불안을 말하기 지연이나 간투사 사용, 반복으로 평가한다는 것이 반영되었다. 4번 문항의 경우, 상대의 말을 듣고 적절하게 조언이나 제안, 거절 등의 대응을 해야 하는 문항이지만 상대방과의 관계나 담화 상황의 성격을 고려하여 내용을 조정하는 능력보다는 조언이나 제안, 거절 등의 내용이 충실히 수행되었는지에 더 초점이 맞추어져 평가되는 특성이 있었다.¹⁰⁾ 4번 문항의 최종 평가 기준은 다음과 같다.

〈표 7〉 4번 문항 최종 평가 기준

평가 요소	내용
내용 및 과제 수행 (5-4-3-2-1-0)	① 과제에 적절한 내용으로 표현하였는가? (과제 수행의 적절성) ② 주어진 과제를 풍부하고 충실하게 수행하였는가?(과제 수행의 풍부성) ③ 담화 구성이 조직적으로 잘 이루어졌는가?
언어사용 (4-3-2-1-0)	① 담화 상황에 적합한 언어를 사용하였는가? (사회문화적 맥락에 대한 이해)-표현의 적절성 ② 어휘와 표현을 다양하고 풍부하게 사용하였는가?(표현의 풍부성) ③ 어휘와 표현을 정확하게 구사하였는가? (표현의 정확성)
발화 전달력 (3-2-1-0)	① 발음과 억양이 어느 정도 이해 가능한가? ② 발화 속도가 자연스러운가?

10) 토픽 말하기 평가가 학습자의 화용 능력까지도 평가할 수 있는 도구로 기능하기 위해서는 평가 기준이 화용 능력을 평가할 수 있어야 한다. 그러나 더 근본적으로는 말하기 평가 문항이 학습자의 화용 능력을 요구하는가를 고찰해보아야 할 것이다. 대화가 아닌 일방적인 말하기 평가 상황에서 상황과 맥락을 고려하여 말하는가를 평가하는 것이 쉽지 않으나 학습자의 의사소통 능력을 평가하기 위한 평가라면 화용 능력까지도 파악할 수 있는 문항을 개발하는 것은 중요한 연구 과제라고 할 수 있다.

3. 1, 2차 채점 결과의 비교

4.2에서 제시한 최종 평가 기준과 루브릭을 평가자 5인에게 제공하고 2차 채점을 실시하였다. 먼저 2번 문항의 1차 채점과 2차 채점의 결과를 비교하여 보면 다음과 같다.

〈표 8〉 2번 문항 1차 채점 결과¹¹⁾

	R1	R2	R3	R4	R5	평균	표준편차
S1	8	7	8	9	8	8.0	0.71
S2	8	7	6	7	6	6.8	0.84
S3	6	6	4	6	4	5.2	1.10
S4	8	9	8	7	9	8.2	0.84
S5	5	3	2	1	2	2.6	1.52
S6	7	7	8	6	8	7.2	0.84
전체평균						6.5	
전체 표준편차							2.17

〈표 9〉 2번 문항 2차 채점 결과

	R1	R2	R3	R4	R5	평균	표준편차
S1	7	7	7	8	8	7.4	0.55
S2	6	9	7	7	6	7.0	1.22
S3	5	5	4	5	4	4.6	0.55
S4	8	9	8	9	9	8.6	0.55
S5	4	4	4	4	3	3.8	0.45
S6	6	7	8	6	6	6.6	0.89
전체평균						6.3	
전체 표준편차							1.81

11) S1,S2...는 학습자1, 학습자2를 의미하고 R1, R2...는 평가자1, 평가자2를 의미한다.

〈표 8〉과 〈표 9〉의 결과를 비교해보면, 전체 학습자 평균이 1차에서는 6.5점이었으나 2차에서는 6.3으로 평균 점수는 약간 하락했으나 전체적인 점수 분포가 좁아졌음을 알 수 있다. 1차 채점 결과에서 전체 표준 점수는 2.17, 2차는 1.81로 평가자 간 점수 차이가 줄어 점수 간 일관성이 높아지고 평가의 안정성이 개선되었음을 알 수 있다. 특히 S1, S3, S4의 표준편차가 유의미하게 감소되었고, 이 중에서도 S5의 경우 표준편차가 1.07이나 감소되어 평가자간 변동성이 크게 축소되었음을 알 수 있다. 이와 같이 같은 학습자에 대해 채점자들이 1차에 비해 2차에서 더 비슷한 점수를 주었다는 것은 평가 기준에 대한 합의 과정에서 평가자 일관성이 상승되었음을 말해주는 것이다.

〈표 10〉 4번 문항 1차 채점 결과

	R1	R2	R3	R4	R5	평균	표준편차
S1	10	10	9	10	8	9.4	0.89
S2	9	9	7	9	6	8.0	1.26
S3	10	7	7	9	4	7.4	2.15
S4	11	11	11	9	8	10.0	1.26
S5	7	9	5	6	4	6.2	1.94
S6	10	11	9	11	7	9.6	1.52
전체평균						8.43	
전체 표준편차							1.58

〈표 11〉 4번 문항 2차 채점 결과

	R1	R2	R3	R4	R5	평균	표준편차
S1	7	10	9	9	12	9.4	1.81
S2	7	9	7	9	6	7.6	1.14

S3	7	10	8	9	7	8.2	1.30
S4	8	9	12	10	12	10.2	1.64
S5	3	6	6	5	4	4.8	1.17
S6	9	12	10	11	12	10.8	1.17
전체평균						8.5	
전체 표준편차							1.43

4번 문항의 1, 2차 채점 결과를 비교해보면 전체 평균은 1차에 비해 2차가 약간 높아졌고 표준 편차는 약간 줄어 점수 분포가 좁아진 것을 알 수 있다. 학습자 간의 점수 차이는 유지되었지만 평가자 간 점수 차이가 다소 줄어들었다. S1, S4, S6은 1, 2차에서 모두 평균을 유지하거나 상승하며 상위권의 안정성을 유지하고 있는데 하위권 점수인 S3, S5의 경우 1차에서 평가자마다 점수 차이가 컸다. 그러나 2차 채점에서는 1차 보다 점수 편차가 줄어들어 평가자 간 일관성이 개선되었음을 확인할 수 있다. 특히 S5처럼 점수가 낮았던 학습자의 편차가 줄어든 것은 평가자 간 판단 기준이 통일되었음을 시사한다.

1차와 2차 채점 후, 평가 기준 및 평가 과정에 대한 의견을 자유롭게 제시하도록 하였다. 채점자들의 의견은 먼저, 평가 기준에 대한 합의의 과정을 통해 기존 말하기 평가 기준에 비해 평가에 용이성이 높아졌다는 것이었다.

- 기존의 루브릭보다 구체적으로 평가 기준이 제시되어 평가하기에 수월했지만, 평가의 점수차가 1점차를 넘지 않아야 한다는 조건이 있어 평가에 제약을 주었다.(R3)
- 각 문항별 구체적인 평가 기준이 마련되어 채점하기에 편리하였고, 평가 요소별 점수가 있어서 평가하기에 용이했다.(R4)
- 점수 배점에 따라 주의해서 들어야 할 내용에 더 집중할 수 있었다.(R5)

반면, 이러한 평가 지침의 통일과 관련된 합의는 계속 이루어지고 평가자 교육이 계속 필요하다는 점에 대한 의견도 있었다.

- 발화 전달력 부분에서 발음과 억양을 함께 평가하도록 되어 있어, 발음은 이해가능하지만 억양이 어색한 경우 채점이 고민된다.(R2)
- 점수 구간의 ± 1 부분의 기준점이 모호하다. 각 평가 요소에 따른 것인지, 아니면 내용 및 과제 수행을 기준으로 하는 것인지를 명확히 할 필요가 있다.(R4)
- 쓰기가 아닌 순간적이고 짧은 말하기를 평가하는 것이 쉽지 않아, 평가자들 역시 많은 연습과 노력이 필요한 듯 하다.(R1)

기존의 말하기 평가 기준은 각 문항의 특성이 다름에도 불구하고 동일한 평가 요소 및 내용으로 평가해야 했기에 실제 토픽 평가자들의 경우 채점자 교육을 통해 각 문항에 대한 채점 매뉴얼을 공유하고 할 것이라 예상할 수 있다. 그러나 현장에서 토픽 말하기를 교육하는 교수자나 토픽 말하기 시험을 준비하는 한국어 학습자에게는 다소 모호한 평가 기준으로 여겨질 수 있다.

그러나 앞서 살펴본 1, 2차 채점 결과 및 전문가 의견을 통해 평가 기준의 재설정 과정과 루브릭을 통해 채점 결과의 일관성이 향상되고, 평가의 용이성이 높아졌음을 알 수 있다. 즉, 2번, 4번 문항의 특성에 맞는 평가 내용의 구체화 및 합의 과정을 통해 평가자에게 유의미한 도움을 주었고, 또한 구체적인 평가 내용을 학습자에게 제공할 경우, 학습의 방향을 명확하게 제시할 수 있는 이점도 있을 것이라고 본다.

V. 결론 및 제언

이 연구는 토픽 말하기 평가 기준을 화용 능력 평가의 관점에서 비판적으로 고찰하고 토픽 말하기 평가가 말하기 화용 능력을 포함한 의사소통 능력 전반을 평가할 수 있는 평가 도구가 될 수 있도록 평가 기준을 재설정하기 위한 것이다. 이를 위해 토픽 말하기 평가 문항 중 상황과 맥락을 고려하여 말하는 능력이 중요하게 요구되는 2, 4번 문항을 연구의 대상으로 삼고 한국어 학습자 6인의 말하기 수행 자료를 수집하였다. 다음으로 한국어 말하기 평가 전문가 집단 5인을 구성하여 1, 2차에 걸쳐 채점을 실시하고 의견을 교환하며 평가 기준을 수정하였다.

1차 채점 후 평가자들은 주로 억양이나 뉘앙스, 사회적 맥락에 맞지 않는 발화 등 화용 능력과 관련된 평가 기준이 부재하다는 의견을 제시하였고, 이를 반영하여 ‘내용 및 과제 수행, 언어 사용’ 요소에서 화용 능력과 관련된 평가 내용을 추가하여 수정하였다. 1차 평가 기준 수정안에 대해 평가자들은 각 항목별 중요도에 따라 배점을 재조정할 것, 말하기 불안이나 긴장에 대한 평가 기능성에 대한 의문을 제기하였다. 또한 평가 기준을 평가에 적용할 때 발생하는 모호함을 구체화시키고자 합의의 과정을 거쳤다. 평가자들이 제시한 의견들과 채점 결과를 바탕으로 토픽 말하기 평가 기준에 대한 재설정 방향을 제시하면 다음과 같다.

첫째, 2번과 4번 문항은 다른 문항에 비해 상황과 맥락을 고려하여 적절한 발화를 할 수 있는 능력이 요구되므로 이러한 평가의 기준이 ‘내용 및 과제 수행’ 요소에 포함되어야 한다. 수정된 2번 문항의 내용 및 과제 수행 요소에는 ‘과제에 적절한 내용으로 풍부하게 표현하였는가, 과제가 상황과 맥락에 맞게 표현되었는가’의 두 가지 내용이 포함되었다. 평가자들은 이러한 과정에서 평가 기준에 대해 명확히 인식할 수 있었고, 그 결과 평가 결과의 신뢰도도 유의미하게 상승하였음을 1차 채점과 2차 채점의 결과 비교를 통

해 확인할 수 있었다.

둘째, 학습자의 발화가 담화 상황에 적절한가를 평가할 때 ‘내용 및 과제 수행’ 요소와 ‘언어 사용’ 요소로 나누어 기준을 구체화시켜 평가자들이 다른 평가 요소로 인식할 수 있도록 해야 한다. 예를 들어 요청을 하는 상황이라면 부탁의 내용이 제대로 포함되어 있는가는 ‘내용 및 과제’ 요소에서 평가하고 부탁을 위해 ‘~아/어 주세요, ~을 수 있을까요?’ 등의 표현을 사용할 경우 ‘언어 사용’ 요소에서 평가할 수 있다는 것을 평가 기준에 구체화 시키는 것이 필요하다는 것이다.

셋째, 발화전달력 요소의 평가 내용이 보다 구체화되어야 한다. 기존 말하기 평가 기준은 ‘발음과 억양’, ‘발화 속도’의 두 가지로 구성되어 있고, 이 연구에서도 두 가지 평가 내용은 유지하되 ‘발화 속도’에서 말하기 지연, 간투사, 반복 등의 내용을 설정하여 함께 평가하도록 합의하였다. 그러나 2차 채점 후의 의견에서 발음은 이해가능하나 억양이 어색한 경우 어떻게 채점할 것인가에 대한 의문을 제기하였다. 평가 내용에 대한 명확한 구분이 없을 경우, 평가 점수가 학습자의 수준을 타당성 있게 설명해주기 어렵다는 문제가 발생할 수 있다.

넷째, 평가 기준을 적용하기 위한 평가 지침에 대한 내용으로 점수의 배점을 평가 요소의 중요성에 따라 설정해야 한다는 것이다. 기존의 말하기 평가에서는 각 문항의 배점이 제시되고 총점은 200점이라고 명시되어 있을 뿐이다. 실제 평가의 현장에서는 평가자들 간에 점수의 배점이 공유될 것이나 각 문항 점수에서 평가 요소에 따라 동일한 점수를 배분하는 것이 아니라 중요도에 따라 점수를 배분하는 것이 타당하다고 본다.

다섯 째, 평가자들의 평가 용이성을 높이고 평가 신뢰도를 높이기 위해 평가 루브릭을 제공하는 것이 필요하다. 말하기 평가는 쓰기 평가에 비해 높은 집중력이 필요하고, 평가자 훈련 및 평가 기준에 대한 합의 과정도 중요하다. 또한 평가 기준을 실제 평가 상황에 적용하기 위한 과정이 필요한데, 평가 루브릭을 통해 평가자들은 학습자 수행에 보다 집중할 수 있는 이점이

있다. 이 연구에 참여한 5인의 평가자가 평가 루브릭을 통해 1차에 비해 2차 평가 시 평가 상황에 집중할 수 있었다는 의견을 제시한 것을 통해 효과를 확인할 수 있었다.

이상의 토픽 말하기 평가 기준의 재설정안은 평가 기준에 대한 세 가지 방향과 두 가지 평가 지침에 대한 것이다. 이 연구는 기존 말하기 평가 기준이 실제 의사소통 상황에서 중요한 능력인 화용 능력까지도 평가할 수 있도록 재설정하기 위해 학습자를 대상으로 말하기 평가를 실시하고 전문가들이 수차례의 합의의 과정을 거쳐 최종 평가 도구 및 평가 루브릭, 평가 도구의 재설정안을 제시하였다는 점에서 향후 토픽 말하기 평가 연구에 기초 자료를 제공할 수 있다고 본다.

다만, 실제 토픽 말하기 평가 기준이 평가 점수에 적용되는 과정과 평가 지침에 대한 구체적인 정보를 파악하는 데에 어려움이 있어 이 연구에서 제시한 재설정안을 실제 토픽 말하기 평가에 적극적으로 반영시키기에는 한계가 있을 수 있으나 이 연구가 교육현장에서 토픽 말하기를 지도하는 교수자와 토픽 말하기 시험을 준비하는 학습자들에게 작은 도움이 될 수 있기를 기대한다.

* 본 논문은 2025.10.31. 투고되었으며, 2025.11.10. 심사가 시작되어 2025.12.03. 심사가 종료되었음.

참고문헌

- 강수정·김영규(2016), 「TOPIK 내용 범주의 재설정을 통한 화용 능력 평가를 설정에 관한 연구」, 『언어정보』 22, 5-40.
- 강현주(2019), 「한국어 말하기 평가 맥락에서 사회언어적 능력 측정 방안 연구」, 『한국어교육』 30(2), 1-23.
- 김도연·난춘야·문보라(2024), 「TOPIK 말하기 평가 ‘그림 보고 이야기하기’ 문항 교육의 방향성 고찰-중국인 학습자를 대상으로-」, 『국제어문』 97, 261-287.
- 김유미(2024), 「ChatGPT 기반 TOPIK 말하기 평가피드백 모형 개발 및 적용」, 『이중언어학』 97, 1-37.
- 성인경ㄱ(2024), 「루브릭을 활용한 한국어 학습자의 화용 능력 평가 연구 -요청화행 평가를 중심으로-」, 『한국어교육』 35(3), 31-58.
- 성인경ㄴ(2024), 「한국어 학습자의 화용 능력 평가를 위한 구인 연구」, 『한국언어문화학』 21(2), 1-26.
- 양세정(2023), 「한국어 능력시험(TOPIK) 말하기 평가 분석과 응시자 피드백 고찰」, 『어문론집』 95, 533-566.
- 위약요·최은경·김성주(2023), 「메타버스를 활용한 TOPIK 말하기 연습 방안 연구-KFL 학습자를 대상으로-」, 『문화와 융합』 45(6), 119-132.
- 이은영(2024), 「한국어 말하기 평가 구인 연구」, 배재대학교 박사학위논문.
- 장지영·박수연(2025), 「한국어 말하기 평가 대비를 위한 생성형 AI의 활용 가능성 탐색-멀티모달(Multimodal)기반 음성 상호작용 기능을 중심으로」, 『교양교육연구』 19(4), 205-216.
- 정민수(2018), 「한국어 학습자의 한국어 화용 능력 평가 분석 연구」, 한국외국어대학교 대학원 석사학위논문.
- 조희영(2024), 「한국어 학습자를 위한 TOPIK 말하기 평가의 대화 완성하기 문항 교육 방안」, 『교사교육연구』 63(4), 526-541.
- 조희영·김지현(2023), 「TOPIK 말하기 교수·학습 방안 연구-5번, 6번 문항에 나타난 학습자의 어려움을 중심으로」, 『어문론집』 96, 505-535.
- Ishihara, N. & Cohen, A. D. (2018), 『다문화 이해를 위한 화용교육』, 조선경·조현용(역), 도서출판 하우(원서출판 2015).

TOPIK 말하기 평가 기준의 재설정을 통한 화용 능력 평가 연구

— 2번 ‘그림보고 역할 수행하기’, 4번 ‘대화 완성하기’를 중심으로

성인경

이 연구는 현재 토픽 말하기 평가 기준을 말하기 화용 능력을 포함한 전반적인 의사소통 능력을 평가할 수 있는 기준으로 재설정하기 위한 것이다. 이를 위해 토픽 2번과 4번 문항을 대상으로 한국어 학습자 6인의 말하기 수행 자료를 수집하고 말하기 평가 전문가 집단을 구성하여 1, 2차에 걸쳐 채점을 실시하였다. 이 과정을 통해 기존 말하기 평가 기준에 대한 의견을 교환하며 최종 평가 기준 및 재설정 방향을 마련하였다.

첫째, 내용 및 과제 수행 요소에 과제가 상황과 맥락에 맞는 내용인가, 상황과 맥락에 적절한 표현인가를 평가하는 내용이 포함되어야 한다는 것 둘째, 답화 상황에 대한 적절성을 평가할 때 ‘내용 및 과제 수행’ 요소와 ‘언어 요소’로 나누어 평가 기준을 보다 구체화시켜야 한다는 것, 셋째, 발화 전달력 요소의 평가 내용에서 발음, 억양, 속도 등의 평가 기준이 구체화되어야 한다는 것이다. 평가 기준을 실제 평가에 적용하기 위한 지침으로 점수의 배점을 평가 요소의 중요성에 따라 설정할 것, 평가 용이성과 신뢰도를 높이기 위한 평가 루브릭을 제공할 것을 제안하였다. 이 연구는 토픽 말하기를 지도하는 교수자들에게 실질적인 도움을 제공하고 향후 토픽 말하기 평가 연구의 기초자료로 사용될 수 있을 것으로 기대한다.

핵심어 TOPIK 말하기 평가, 화용 능력, 화용 능력 평가 기준, 평가 기준 재설정

A Study on the Assessment of Pragmatic Competence through the Redefinition of the TOPIK Speaking Evaluation Criteria

— Focusing on Task 2 “Role-Play Based on a Picture” and Task 4 “Completing a Dialogue”

Seong Inkyoung

This study aims to redefine the current TOPIK speaking evaluation criteria so that they can assess not only linguistic accuracy but also overall communicative competence, including pragmatic competence. To this end, speaking performance data were collected from six Korean language learners based on Tasks 2 and 4 of the TOPIK Speaking Test. A group of speaking assessment specialists was organized to conduct two rounds of rating. Through this process, the evaluators exchanged opinions on the existing criteria and developed final evaluation standards and directions for revision.

First, it was concluded that the “content and task performance” category should include an assessment of whether the response content and expressions are appropriate to the given situation and context, thereby reflecting the learner’s pragmatic competence. Second, in evaluating the appropriateness of discourse situations, the criteria should be further specified by dividing them into two elements: *content and task performance* and *linguistic resources*. Third, in the delivery category, the evaluation standards for pronunciation, intonation, and speech rate should be made more detailed and explicit. As practical guidelines for applying the proposed criteria to actual assessments, the study proposes determining the weighting of each scoring category according to its relative importance and using a detailed evaluation rubric to enhance scoring feasibility.

and reliability. It is expected that this study will provide practical insights for instructors teaching TOPIK speaking and serve as foundational data for future research on the assessment of pragmatic competence in TOPIK speaking.

KEYWORDS TOPIK Speaking Test, pragmatic competence, evaluation criteria for pragmatic competence, redefinition of evaluation criteria